

## Statistische Analysen mit SAS und LogXact

Dr. Andreas Christmann

Universität Dortmund, Hochschulrechenzentrum

### Abstract

SAS ermöglicht die Anpassung logistischer Regressionsmodelle auf verschiedene Arten. Insbesondere bei kleinen Fallzahlen oder bei sehr asymmetrischen Designs bietet sich die Verwendung eines bedingten Maximum Likelihood Schätzers bzw. der bedingten exakten logistischen Regressionsschätzung an. Meines Wissens sind derartige Analysen derzeit effizient nur in der Software LogXact von CYTEL verfügbar. In dem Vortrag wird eine einfache Methode beschrieben, wie man Daten aus SAS nach LogXact übergeben, dort die exakten Analysen durchführen und die Ergebnisse wieder nach SAS importieren kann, ohne das SAS-System verlassen zu müssen. Diese einfache Methode hat sich in einem statistischem Beratungsfall als sehr nützlich erwiesen, da dort viele kleine Datensätze vergleichbaren Typs zu analysieren waren.

### Einleitung

Die logistische Regression ist ein klassisches Verfahren, um Ereignis-Wahrscheinlichkeiten zu modellieren und wird in vielen Bereichen verwendet. Die Zielgrößen  $Y_i$ ,  $i=1, \dots, n$ , werden als stochastisch unabhängig und als Binomial-verteilt  $Bi(1, \Lambda(x_i^T \theta))$  angenommen. Hierbei bezeichnet  $\theta \in \mathbb{R}^p$  den unbekannt Parameter-Vektor und  $x_i = (x_{i,1}, \dots, x_{i,p})^T$  den Vektor der erklärenden Variablen. In der Medizin können so zum Beispiel Erkrankungswahrscheinlichkeiten in Abhängigkeit von Risikofaktoren modelliert werden. Ziel ist die Schätzung und die Durchführung von Tests für  $\theta$ .

SAS ermöglicht die Anpassung logistischer Regressionsmodelle auf verschiedene Arten. Insbesondere sind hierbei die Prozeduren LOGISTIC, GENMOD, PROBIT und CATMOD sowie SAS/INSIGHT zu nennen. Im Gegensatz zu vielen anderen Software-Produkten aus dem Statistikbereich bietet SAS mit der Prozedur PHREG zudem die Möglichkeit, bedingte logistische Regression bei m:n Matching durchzuführen. Die unbekannt Parameter werden im allgemeinen mit der Maximum Likelihood Methode oder mit der WLS Methode geschätzt. Standardfehler und Tests basieren i.a. auf der Annahme einer approximativen Normalverteilung für große Fallzahlen.

### Problemstellung

Insbesondere bei kleinen Fallzahlen oder bei sehr asymmetrischen Designs ist diese Annahme einer approximativen Normalverteilung der klassischen Parameterschätzer jedoch häufig zumindest zweifelhaft. In diesen Fällen bietet sich die Verwendung eines bedingten Maximum Likelihood Schätzers bzw. der bedingten exakten logistischen Regressionsschätzung an. Meines Wissens sind derartige Analysen derzeit effizient nur mit der Software LogXact von CYTEL durchführbar.

Unter Verwendung des „x“-Befehls ist es möglich, SAS-Makros zu schreiben, die die auszuwertenden Daten von SAS an LogXact übergeben und die exakte logistische Regressionsschätzung inklusive Tests mit LogXact durchführen. Die Ergebnisse können dann in SAS eingelesen und gegebenenfalls weiterverarbeitet werden.

## Lösungsansatz

Innerhalb des SAS-Systems wird zunächst eine SAS-Transport-Datei erstellt, um den Datenimport nach LogXact zu ermöglichen. Zudem wird eine ASCII-Datei mit den LogXact-Befehlen zur Durchführung der gewünschten statistischen Analysen erstellt. Danach wird LogXact aus dem SAS-System aufgerufen mit dem Befehl der Form

```
x 'd:\logxact\logxact.exe d:\tmp\syntax.cmd';
```

Mit der Taste F10 werden die Analysen durch LogXact gestartet. Danach wird der LogXact-Output als ASCII-Datei abgespeichert. Damit ist das Importieren der relevanten Ergebnisse in SAS möglich, um diese dort weiter zu verarbeiten.

Diese einfache Methode hat sich in einem statistischem Beratungsfall als sehr nützlich erwiesen, da dort viele kleine Datensätze vergleichbaren Typs zu analysieren waren und die Verwendung einer multiplen Testprozedur notwendig war. Mit der SAS-Prozedur MULTTEST konnten die einzelnen p-Werte der mit LogXact durchgeführten Signifikanztests mit der Bonferroni-Holm-Methode adjustiert werden.

## Beispiel

Das folgende SAS-Programm illustriert die wesentlichen Programmierschritte, wobei keinerlei Gewähr übernommen wird.

```
*****;
* SAS/Macro fuer exakte logistische Regression mit Cytel/LogXact *;
* SAS fuer Windows 6.12, Cytel/LogXact 2.0 *;
* KEINERLEI GEWAHR *;
* dat Datensatz *;
* y Responsevariable (0/1) *;
* model Modell, z.B. x1+x2 (Wichtig: "+") *;
* estim gewuenschte Parameterschaetzer, z.B. x1 x2 (ohne "+") *;
* test gewuenschte Tests, z.B. x1 (ohne "+") *;
* strat Stratifizierungsvariable fuer bedingte logistische *;
* Regression *;
* estim, test, strat = . wenn nicht gewuenscht *;
*****;

%MACRO sas2loxt(dat,y,model,estim,test,strat);
  OPTIONS MPRINT;
  %LET dat=&dat; %LET y=&y;
  %LET model=&model; %LET estim=&estim;
  %LET test=&test; %LET strat=&strat;
  *** Transport-Datei erstellen;
  LIBNAME out XPORT "d:\tmp\temptmp.xpt";
  DATA out.tmp; SET &dat; RUN;
```

```
*** ASCII-Datei mit LogXact-Befehlen erstellen;
DATA _NULL_;
  FILE 'd:\tmp\syntax.cmd';
  LENGTH text $ 80;
  %LET text="IM SAS d:\tmp\tmpmp.xpt"; PUT &text;
  u=SYMGET('y'); text='RE '; PUT text u;
  %IF &strat NE . %THEN %DO;
    u=SYMGET('strat'); text='ST '; PUT text u; %END;
  u=SYMGET('model'); text='MO '; PUT text u;
  %IF &estim NE . %THEN %DO;
    u=SYMGET('estim'); text='ES EX %Const'; PUT text u; %END;
  %IF &test NE . AND &strat EQ . %THEN %DO;
    u=SYMGET('test'); text='TE EX %Const'; PUT text u; %END;
  %IF &test NE . AND &strat NE . %THEN %DO;
    u=SYMGET('test'); text='TE EX '; PUT text u; %END;
  %LET text="QU"; PUT &text;
run;
*** LogXact aufrufen und die Analysen durchführen;
  x 'd:\logxact\logxact.exe d:\tmp\syntax.cmd';
RUN;
%MEND;

DATA a;
  INFILE ...;
  INPUT x1 x2 y;
RUN;
%sas2loxt(a, y, x1+x2, x1, ., .);
*** LogXact-Output einlesen;
DATA out1;
  INFILE ...;
  INPUT ... test$ ... raw_p ...;
RUN;
*** Durchführung der multiplen Testprozedur;
PROC MULTTEST PDATA=out1 HOLM OUT=out2;
PROC PRINT DATA=out2;
RUN;
```

## Literatur

- Agresti, A. (1996). An Introduction to categorical data analysis. Wiley & Sons, New York.
- Cox, D.R., Snell, E.J. (1989). Analysis of binary data. 2<sup>nd</sup> edition. Chapman & Hall, London.
- LogXact for Windows (1996). CYTEL Software Corporation, Cambridge, USA.
- SAS Institute Inc., SAS/STAT User's Guide, Version 6, Fourth Edition, Volume 1, Cary, NC: SAS Institute Inc., 1989.
- SAS Institute Inc., SAS/STAT User's Guide, Version 6, Fourth Edition, Volume 2, Cary, NC: SAS Institute Inc., 1989.
- SAS Institute Inc., SAS/STAT Software: Changes and Enhancements through Release 6.12, Cary, NC: SAS Institute Inc., 1996.